2014 STI conference – 3-5 September 2014 – Leiden, The Netherlands

Differences between examiner and applicant citations in the European Patent Office: a first approach

Joaquín M. Azagra-Caro and Elena M. Tur*

* *jazagra@ingenio.upv.es*, *elmatu@ingenio.upv.es*

INGENIO (CSIC-UPV), Ciudad Politécnica de la Innovación Edif. 8E 4º, Camino de Vera s/n, 46022 Valencia

Abstract–In the US Patent Office, examiners add extra shares of citations to foreign applicants. We explore a similar country club effect in the European Patent Office (EPO). Using EPO data of over 3,500,000 citations in years 1997-2007, we find national variation in the probability of an applicant originating a citation rather than the examiner. Symmetrically to the US case, EPO examiners add extra citations to non-signatory member states. Moreover, if examiners are likely to come from the same country of the applicants, applicant-citation shares increase, pointing to the existence of national bias in EPO patent examiners. These results hold after controlling for sub-national characteristics of the patenting process.

Keywords–Citations, knowledge flows, national biases

Introduction

The geography of innovation makes extensive use of backward citations in patents to measure knowledge flows (Jaffe et al., 1993). Several works emphasize the importance of distinguishing the origins of citations because, in theory, citations inserted by patent examiners are likely to be less localized than applicant citations. United States Patent and Trademark Office (USPTO) data mostly tend to confirm this for the US case (Thompson, 2006) although there are some differences for some specific measures of distance (Alcácer & Gittelman, 2006). European data confirm it for some European countries (Criscuolo & Verspagen, 2008), but not some regions with low absorptive capacity (Azagra-Caro et al., 2009). These studies focus on the match or distance between citing and cited country. However there is another geographic concern that has been largely unexplored, i.e. what are the characteristics of the citing country? Do patent examiners add more citations to patent applications from specific countries?

This is a relevant question because the answer might reveal underlying economic forces that are subject to policy influence, or uncover individual questionable examiner practices. There is some evidence suggesting that, for the USPTO, geographic origin of the applicant matters, e.g. US examiners add more citations to foreign applications (Alcácer et al., 2009). However, there is a lack of research on a similar 'club effect' in the case of the European Patent Office (EPO). This is unfortunate because the EPO is frequently used as a benchmark against the USPTO, and is considered one of the highest quality patent systems due to its rigorous granting process and flexibility applied to later stages in a patent's life (Saint-George & van Pottelsberghe, 2013). This paper focuses on the EPO. By comparing with the USPTO, we should be able to identify whether there is a symmetrical geographical effect, namely whether EPO patent examiners are more likely to add citations to foreign applications: Do EPO examiners add extra citations to applications from countries outside the European Patent Organization (EPOrg)? And do EPO examiners add extra citations to applications from countries other than their own?

Model, data and variables

We estimate the following model:

$$\Pr(appcit_{ijklt}) = f(\alpha X_{it}, \beta X_{jt}, \gamma X_{kt}, \delta X_{lt}) + \varepsilon_{ijklt} \quad (1)$$

where *appcit* is equal to 1 if the citation is inserted by the applicant and 0 if inserted by the examiner. The probability varies according to the characteristics of the citation *i*, the patent *j*, the applicant *k* and the applicant country *l*. The year of the patent application *t*, is lagged two periods for national economic and research and development (R&D) characteristics to prevent endogeneity.

Data on patents and citations come from Patstat (October 2012 edition). We selected patents where the publication authority was the EPO –almost 2.5 million. After removing those with missing or unreliable information for application year and technology class (represented by the International Patent Classification IPC), and those without citations, we were left with 2 million patents.

Those patents contained over 12 million citations. Patstat classifies them into origin types, i.e. the moment in the examination process when the citation was inserted. There are ten types of origins (coded 0-9), but only some are relevant for this study, i.e. those indicating that either patent applicant or examiner could have inserted the citation (see section **¡Error! No se encuentra el origen de la referencia.** for further details): origins coded 0 (citations introduced during search), 2 (citations introduced during examination) and 5 (citations from the International Search Report). They represent most (82%) of the citations.

Patstat differentiates who inserted the citation by classifying citations with origins 0, 2 and 5 into several categories. Categories (coded with single letters, A, X, Y, etc.), refer to the relevance of prior art to invalidate claims of novelty. Criscuolo & Verspagen (2008) call category D 'applicant citations' and sum the other categories as 'examiner citations'. We follow this method.

In the estimations, the number of observations is not the number of citations for two reasons. First, duplicates are created if the patent has more than one applicant. We deal with this econometrically by weighting the observations by the inverse number of applicants. Second, we match Patstat to other databases on national characteristics that do not have full information for all countries and years. The sample includes over 3.6 million observations. The proportion of D-citations in the total is our dependent variable, computable for over 7 million citations.

Table 1 provides information on the econometric model variables.

The dependent variable is a dummy that takes the value 1 if the citation comes from the examiner. A logit model is appropriate for this kind of data.

Table 1. Variable definitions and descriptive statistics (n=3,663,276)

Vector	Name	Source	Variables	Description	Mean	Std. Dev.	Min	Max
$appcit_{ijklt}$	Applicant citation	Patstat	Citation category D	1 if citation category is D, 0 if other category	0.07	0.26	0.00	1.00
X_{it}	Citation characteristics	Patstat	Non-patent literature	1 if non-patent literature, 0 if patent literature	0.36	0.48	0.00	1.00
			European search report	1 if origin in search report	0.85	0.36	0.00	1.00
			Examination report	1 if origin in examination	0.00	0.06	0.00	1.00
X_{jt}	Patent characteristics	Patstat	Euro-PCT	1 if EPO-PCT, 0 if direct EPO	0.46	0.50	0.00	1.00
			Grant	1 if granted, 0 otherwise	0.18	0.39	0.00	1.00
			Filing year	Application year	2001.94	3.03	1997.00	2007.00
			A Human Necessities	1 if IPC code is A Human Necessities	0.21	0.41	0.00	1.00
			B Performing Operations; Transporting	B Performing Operations; Transporting	0.26	0.44	0.00	1.00
			C Chemistry; Metallurgy	C Chemistry; Metallurgy	0.22	0.41	0.00	1.00
			D Textiles; Paper	D Textiles; Paper	0.02	0.14	0.00	1.00
			E Fixed Constructions	E Fixed Constructions	0.04	0.19	0.00	1.00
			F Mechanical Engineering; Lighting; Heating; Weapons; Blasting	F Mechanical Engineering; Lighting; Heating; Weapons; Blasting	0.14	0.34	0.00	1.00
			G Physics	G Physics	0.26	0.44	0.00	1.00
			H Electricity	H Electricity	0.25	0.43	0.00	1.00
X_{kt}	Applicant characteristics	ECOOM*	Individual	1 if institutional sector is… Individual only	0.08	0.26	0.00	1.00
			Government	Government only	0.03	0.16	0.00	1.00
			University	University only	0.02	0.15	0.00	1.00
			Hospital	Hospital only	0.00	0.04	0.00	1.00
			Company-government	Company and government	0.00	0.04	0.00	1.00
			Company-university	Company and university	0.00	0.00	0.00	1.00
			Company-hospital	Company and hospital	0.00	0.01	0.00	1.00
			Government-university	Government and university	0.00	0.00	0.00	1.00
			# applications	Number of applications (millions)	0.00	0.00	0.00	0.15
X_{lt}	Country of applicant characteristics – economic and R&D	OECD R&D Statistics	GDP	Real Gross Domestic Product (GDP): billion Euro	0.04	0.04	0.00	0.13
			GDP per capita	GDP: Euro per inhabitant (millions)	0.03	0.01	0.00	0.07
			GERD intensity	Total intramural Gross R&D expenditure (GERD): Millions of Purchasing Power Standards (PPS) at 2000 prices	2.51	0.47	0.28	4.58
			% business funding of R&D	Business R&D funding: Share of GERD	0.64	0.09	0.17	0.91
	Country of applicant characteristics – related to EPO	EPO Annual Reports	Prob EPO exam same country	Probability of examiner from same nationality	0.10	0.10	0.00	0.26
			EPOrg member	EPO member (yes/no)	0.44	0.50	0.00	1.00

* Methodology for construction of ECOOM data explained in DuPlessis et al. (2009), Magerman et al. (2009) and Peeters et al. (2009).

Results

0 presents the estimations. Column 1 includes the specification of Equation 1 with citation and patent characteristics only; the remaining columns include the variables progressively.

Citation and patent characteristics

The results for the sub-national variables are consistent across estimations. Citations are coded to indicate whether the origin is a Euro-PCT (not a direct EPO) application, and whether it is the European search report or the examiner report (rather than the international search report). The coefficient of "Euro-PCT" is negative and significant, indicating that this longer procedure leads to higher numbers of examiner citations. The coefficient of "European search report" is negative and significant, implying that citations in this second phase are more likely to be associated with examiners than if there was an international search report in the first phase. The coefficient of "Examiner report" is also negative and significant and higher than the coefficient of "European search report", meaning that citations in this third phase are most likely to come from examiners.

The sample includes applications and grants. This is controlled for in the models by the dummy variable "Grant". The estimated coefficient is positive and significant. Hence, we can confirm a link between receiving relatively fewer examiner citations and having the patent granted. In part, this is intuitive. It becomes more interesting if we consider that, in the USPTO, this does not necessarily apply. In the USPTO, more experienced examiners, and examiners that systematically cite less prior art, are more likely to award patent grants (Lemley & Sampat, 2012). Moreover, USPTO examiners rarely use applicant citations to reject a grant (Cotropia et al., 2013). Hence, examiner citation shares are not associated with denial of a grant in the USPTO but they are in the EPO. This and other signs may indicate the superiority of the EPO patent system (Saint-George & van Pottelsberghe, 2013).

We test whether applicants are more likely than examiners to cite non-patent literature, extrapolating from US evidence that examiners rarely cite non-patent literature (Sampat, 2004). The positive and significant sign of "Non-patent literature" shows that this is the case. Applicants are probably more familiar with the fundamental knowledge base underpinning their inventions, while examiners are often engineers whose expertise is related more to parcels of applied knowledge.

Applicant characteristics

Dummies for organizational type of the applicant (models 2-3) can be used to validate empirically which one matters more. "Company only" is the benchmark. The positive, significant coefficients of "Government only" and "University only" indicate that these institutions generate more reliability than corporate patents. The coefficients of "Individuals only" and "Hospital only" are negative and significant, which means that citations are less likely to originate in applicants than in the case of firms. Individuals may show lower citation shares because institutions facilitate settings where citing is more common practice, i.e. through sharing of references and codified knowledge. Examiner citation shares may be larger for hospitals because they do not have a

tradition of patenting, and on patents related to clinical practice which are less related to science.

Table 2. Logistic regression of the probability of an applicant originating a citation rather than the examiner

	1 Citation and patent characteristics	2 + Applicant characteristics	3 + Country characteristics
Euro-PCT	-0.68***	-0.68***	-0.48***
	(0.01)	(0.01)	(0.01)
European search report	-0.93***	-0.94***	-0.57***
	(0.01)	(0.01)	(0.01)
Examination report	-2.73***	-2.74***	-2.43***
	(0.09)	(0.09)	(0.09)
Grant	0.30***	0.29***	0.29***
	(0.00)	(0.00)	(0.00)
Non-patent literature	0.06***	0.05***	0.10***
	(0.01)	(0.01)	(0.01)
Individual		-0.15***	-0.21***
		(0.01)	(0.01)
University		0.04***	0.08***
		(0.01)	(0.01)
Government		0.13***	0.05***
		(0.01)	(0.01)
Hospital		-0.39***	-0.31***
		(0.07)	(0.07)
Company-government		-0.09*	-0.10*
		(0.05)	(0.05)
Company-university		1.16***	1.14***
		(0.29)	(0.30)
Company-hospital		0.48*	0.31
		(0.27)	(0.27)
Government-university		-0.17	-0.40
		(0.52)	(0.53)
# applications		-0.91	-10.64***
		(0.56)	(0.57)
GDP			0.84***
			(0.14)
Per capita GDP			18.77***
			(0.86)
GERD intensity			0.27***
			(0.01)
% business funding of R&D			-0.88***
			(0.04)
Prob EPO exam same country			0.61***
			(0.04)
EPOrg member			0.64***
			(0.01)
Constant	37.53***	38.74***	67.19***
	(1.44)	(1.45)	(1.64)
Observations	3,663,276	3,663,276	3,663,276
Log likelihood	-848,023	-847,774	-838,745
χ^2	54,181	54,658	75,414
Prob>χ^2	0.000	0.000	0.000

* p<0.1; ** p<0.05; *** p<0.01. Robust standard errors in parenthesis. No collinearity according to Variance Inflation Factors. All models include a trend and eight IPC section dummies. Weight: share of number of applicant countries.

Models 2-3 include dummies for types of organizational interactions (taking "Company only" as benchmark). University-company co-applications for patents are strongly associated with a higher probability of an applicant rather than the examiner including a citation. Somewhat surprisingly, government-company co-application for patents is negatively related to that probability. A possible reason might be that organizations in the category government have heterogeneous missions. Government labs with an industry orientation are more likely to engage in partnerships with firms that lead to patents, than labs with an academic orientation, and the government-company dummy captures this type of partnership. This double industry orientation receives a higher share of examiner citations. For other interactions ("Company-hospital", "University-government") the dependent variable does not change significantly.

The number of applicant citations decreases with the increase in the number of applications. Alcácer et al. (2009) found the same in the USPTO case. Their explanation is that large applicants prefer "broad patent portfolios, with relatively low value placed on any single invention" (p. 426). Alternatively, it might be that applicants include unrelated cites after the invention or omit relevant cites for strategic reasons (Breschi & Lissoni, 2005). Perhaps experienced applicants learn how to "cheat", and hide a higher number of relevant references.

National characteristics

The variables GDP, per capita GDP and GERD intensity test the assumption that larger, wealthier and scientifically stronger countries are more likely to create conditions favorable to the appearance of novelty. Their positive, significant coefficients provide evidence to support it. Hence, we observe that countries with these favorable endowments benefit from lower examiner citation shares.

The coefficient of the share of business funding variable in model 3 is negative and significant, supporting this expectation. Examiner citation shares are higher in patents from national contexts where the research orientation is towards more applied research.

Country block effects may also play a role in the model. Specifically, we are interested in whether there is a club effect similar to the one shown by Alcácer et al. (2009) in the USPTO case: US applicants receive fewer examiner citation shares than non-US ones. In our EPO sample, this club effect would not be strictly national since the EPO is international. Instead, we propose that such an effect might be visible for countries belonging to the EPOrg. In the model, the dummy is equal to 1 if the applicant country belongs to EPOrg, to capture this phenomenon. The estimation (positive and significant) verifies that there is a lower propensity for EPOrg member states to receive cites from the examiner. Hence, the EPO is similar to the USPTO: outsiders are less warmly received.

Having isolated a club effect, the nationality of examiners might be influential. Collins & Wyatt (1988) detected national chauvinism in citations to non-patent literature in US genetics patents: "it appears that every country is its own best citer" (p.73). However, Meyer (2000) finds no signs of national chauvinism in nanotechnology patent applications to the USPTO from Swedish applicants. In our estimation, the positive, significant coefficient of the probability of an application being examined by an examiner from the same country as the patent applicant provides support for the national bias assumption.

Conclusions

The literature on the geography of knowledge flows has shown that the probability of an applicant rather than the examiner originating a citation depends on differences between citing and cited countries. Our contribution to this stream of literature is that the conditions of the citing country also matter to predict that probability. Our findings show that better national economic and scientific endowments increase applicant citation shares, whereas higher proportions of business funding of R&D foster examiner citation shares. Future research could test which group of determinants (citing country characteristics or citing-cited country differences) matter more.

Previous analyses of the characteristics of applicant versus examiner citation shares found differences across patent and applicant. We show the presence of additional disparities across citation characteristics, namely procedural aspects of the patenting process and knowledge base of the patent. Our results for procedural aspects increase our understanding of the generation of citations in the various phases of the life of an EPO application. Our results for knowledge base suggest the importance of science to provide credibility to applications.

The use of a sample based on EPO applications allowed comparison with earlier works exploiting USPTO evidence. It suggests that large applicant citation shares are more clearly associated with being awarded a patent by the EPO than the USPTO. It also signals that there are similar club effects, which favor EPOrg members at the EPO and US residents at the USPTO. Since the methods used by Alcácer et al. (2009) and those applied in this study differ, interpretation of this comparison should be cautious. A possible avenue of further inquiry could be designing an experiment to enable direct comparison between both data sources.

References

Alcácer, J., and M. Gittelman, "Patent citations as a measurement of knowledge flows: The influence of examiner citations," Review of Economics and Statistics 88:4 (2006), 774-779.

Alcácer, J., M. Gittelman, and B.N. Sampat, "Applicant and examiner citations in U.S. patents: An overview and analysis," Research Policy 38 (2009), 415-427.

Azagra-Caro, J.M., I. Fernández de Lucio, F. Perruchas, and P. Mattsson, "What do patent examiner inserted citations indicate for a region with low absorptive capacity?" Scientometrics 80:2 (2009), 443–457.

Azagra-Caro, J.M., P. Mattsson, and F. Perruchas, "Smoothing the lies: The distinctive effects of patent characteristics on examiner and applicant citations," Journal of the American Society for Information Science and Technology 62:9 (2011), 1727-1740.

Breschi, S., and F. Lissoni, "Knowledge networks from patent data," In H. F. Moed, W. Glänzel and U. Schmoch (Eds.), Handbook of quantitative science and technology research (Netherlands: Kluwer Acad. Publ.. 2005).

Collins, P., and S. Wyatt, "Citations in patents to the basic research literature," Research Policy 17 (1988), 65-74.

Cotropia, C.A., M.A. Lemley, and B.N. Sampat, "Do applicant patent citations matter?" Research Policy 42:4 (2013), 844-854.

Criscuolo, P., and B. Verspagen, "Does it matter where patent citations come from? Inventor vs. examiner citations in European patents," Research Policy 37 (2008), 1892-1908.

Du Plessis, M., B. van Looy, X. Song, and T. Magerman, "Data Production Methods for Harmonized Patent Indicators: Assignee sector allocation," EUROSTAT Working Paper and Studies (2009), Luxembourg.

Jaffe, A. B., M. Trajtenberg, and R. Henderson, "Geographic localization of knowledge spillovers as evidenced by patent citations," Quarterly Journal of Economics 108:3 (1993), 577-598.

Lemley, M., and B.N. Sampat, "Examiner Characteristics and Patent Office Outcomes," Review of Economics and Statistics 94:3 (2012), 817-827.

Magerman T, J. Grouwels, X. Song, and B. van Looy, "Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization," EUROSTAT Working Paper and Studies (2009), Luxembourg.

Meyer, M., "Does science push technology? Patents citing scientific literature," Research Policy 29 (2000), 409-434.

Peeters B., X. Song, J. Callaert, J. Grouwels, and B. van Looy, "Harmonizing harmonized patentee names: an exploratory assessment of top patentees," EUROSTAT working paper and Studies (2009), Luxembourg.

Saint-George, M., and B. van Pottelsberghe de la Potterie, "A quality index for patent systems," Research Policy 42 (2013), 704– 719.

Sampat, B.. "Examining Patent Examination: An Analysis of Examiner and Applicant Generated Prior Art, Working Paper (2004), School of Public Policy, Georgia Institute of Technology.

Thompson, P. "Patent citations and the geography of knowledge spillovers: Evidence from inventor- and examiner-added citations," Review of Economics and Statistics 88:2 (2006), 383-388.